

RESEARCH ARTICLE

Open Access

Transcripts with *in silico* predicted RNA structure are enriched everywhere in the mouse brain

Stefan E Seemann^{1,2}, Susan M Sunkin³, Michael J Hawrylycz³, Walter L Ruzzo^{1,4,5} and Jan Gorodkin^{1,2*}

Abstract

Background: Post-transcriptional control of gene expression is mostly conducted by specific elements in untranslated regions (UTRs) of mRNAs, in collaboration with specific binding proteins and RNAs. In several well characterized cases, these RNA elements are known to form stable secondary structures. RNA secondary structures also may have major functional implications for long noncoding RNAs (lncRNAs). Recent transcriptional data has indicated the importance of lncRNAs in brain development and function. However, no methodical efforts to investigate this have been undertaken. Here, we aim to systematically analyze the potential for RNA structure in brain-expressed transcripts.

Results: By comprehensive spatial expression analysis of the adult mouse *in situ* hybridization data of the Allen Mouse Brain Atlas, we show that transcripts (coding as well as non-coding) associated with *in silico* predicted structured probes are highly and significantly enriched in almost all analyzed brain regions. Functional implications of these RNA structures and their role in the brain are discussed in detail along with specific examples. We observe that mRNAs with a structure prediction in their UTRs are enriched for binding, transport and localization gene ontology categories. In addition, after manual examination we observe agreement between RNA binding protein interaction sites near the 3' UTR structures and correlated expression patterns.

Conclusions: Our results show a potential use for RNA structures in expressed coding as well as noncoding transcripts in the adult mouse brain, and describe the role of structured RNAs in the context of intracellular signaling pathways and regulatory networks. Based on this data we hypothesize that RNA structure is widely involved in transcriptional and translational regulatory mechanisms in the brain and ultimately plays a role in brain function.

Background

In neurons, RNA molecules often have to travel long distances between transcriptional origin (nucleus) and functional destination (axon, synapses, dendrites). Dendrites contain thousands of postsynaptic sites and long-lasting forms of activity-dependent synaptic modifications (memory storage) are believed to require local protein synthesis. Local protein translation implies that mRNAs are transported from the nucleus and localized to dendrites and synapses [1,2]. It has been speculated that RNA secondary structures in mRNA untranslated regions

(UTRs) are involved in these processes [3,4]. In addition, numerous noncoding RNAs (ncRNAs) are expressed in brain [5,6] and mounting evidence indicates important contributions of ncRNAs in brain functions such as memory formation and maintenance [7,8], as well as a host of other functions in mammalian cells. This study further explores these connections by combining the large scale *in situ* hybridization data of the Allen Mouse Brain Atlas (<http://mouse.brain-map.org>) [9] with *in silico* predictions of conserved RNA secondary structure, revealing extensive enrichment of such structures in the adult mouse brain transcriptome.

Post-transcriptional regulation of RNA splicing, editing, transport, stability, localization and translation through UTR signals plays an important role in controlling gene expression. Important examples of stable RNA secondary

*Correspondence: gorodkin@rth.dk

¹Center for non-coding RNA in Technology and Health, University of Copenhagen, Frederiksberg C, Denmark

²IBHV, University of Copenhagen, Frederiksberg C, Denmark

Full list of author information is available at the end of the article

structures are known in both 5' UTRs [10] and 3' UTRs. For instance, the 84-nucleotide (nt) long structure-anchored repression element (CAESAR) in *CTGF* [4] is highly conserved in structure but not in sequence, and is suspected to inhibit translation and affect mRNA stability [11]. Other structural mRNA elements, such as the selenocysteine insertion sequence (SECIS) element and nanos 3' UTR TCE, are targets of RNA binding proteins. Stem-loop structures in untranslated regions are sometimes critical for proper mRNA localization [12-14], such as translocation of the *MAPT* mRNA along axonal microtubules [15] and *ASH1* mRNA to the cortical actin cytoskeleton [16]. RNA binding proteins might localize the *RAB1A* mRNA to specific cytoplasmic regions through recognition of its highly conserved 3' UTR sequence and structure, so that translation would occur close to the location of the respective protein regulating intracellular vesicle transport [17]. A predicted stable RNA structure overlaps the RNA localization region in the 3' UTR of the mRNA encoding myelin basic protein MBP. The structure (but not the sequence) is conserved in human, mouse and rat [18]. The highest affinity site of the RNA-binding protein *Qki* is located within the RNA localization region of *MBP*, suggesting a possible role for *Qki* in restricting *MBP* mRNA to the myelin membrane [19]. In a very distinct manner, many 3' UTRs in mouse are reported to be expressed separately from their mRNAs in a developmentally regulated manner [20], and some reported regulatory mutations in 3' UTRs do not appear to act *in cis* to regulate the expression of the associated mRNA. Some structured 3' UTRs may, thus, act *in trans* as ncRNAs [21].

Long noncoding RNAs (lncRNAs) have recently received increased attention due to their functional diversity in basic molecular and cellular biology [21-24]. In particular, they appear to be deeply entwined with cellular regulatory machinery, both as targets of important transcription factors [25], and as direct cis- and trans-regulators of gene expression through interactions with transcription factors or as indirect regulators through an RNA-binding protein intermediate (transcription factor co-regulators) [26]. Furthermore, they have demonstrated roles in regulation of dosage compensation, imprinting, chromatin state, and epigenetic inheritance by DNA methylation [26]. A hallmark of many small ncRNAs is the critical role of RNA secondary (and tertiary) structure. RNA structure also may have major functional implications for lncRNAs as shown, *e.g.*, for the noncoding co-factor MEG3 of the tumor repressor *p53* [27], and the *p53* regulated transcriptional repressor *lincRNA-p21*, which is tethered to *hnRNP-K* for its proper localization [28].

Several genome-scale screens for stable, conserved RNA secondary structures have found known RNA families and many potentially novel ncRNAs (EVOfold [29],

RNAz [30], CMfinder [31], FOLDALIGN [32]), albeit with significant false discovery rates [33,34]. Classical transfer RNAs, ribosomal RNAs, some microRNAs and many other functional ncRNAs have a weakly conserved sequence, and instead, have a highly conserved functional secondary structure. Hence, comparative analyses that focus on sequence conservation and ignore potential conservation of secondary structure underestimate ncRNA prevalence. Here, we apply CMfinder to search for RNA secondary structures. The method attempts to create structurally optimal alignments from unaligned orthologous input sequences using an expectation-maximization algorithm. Both thermodynamic energies and evidence for conservation of secondary structure, *e.g.* presence of compensatory mutations in putative helices, are part of the evaluation criteria. An appropriate background model distinguishes between significant RNA structures, *e.g.* putative ncRNAs, and structured background.

The key question addressed in this paper is the extent to which RNA structures, both in noncoding transcripts and the UTRs of protein coding transcripts, play biologically important roles in the brain. We address this question by analyzing transcripts expressed in the adult mouse brain as cataloged in the Allen Mouse Brain Atlas (Atlas) for their potential to contain RNA structures predicted by CMfinder. There are Atlas probes for approximately 20,000 RNA transcripts in the adult mouse brain, visualized at cellular resolution by *in situ* hybridization (ISH) [9]. Of these transcripts, 16,900 exhibit cellular expression above background in the adult mouse brain [9]. Expression data within the ISH images is identified and mapped to defined regions [35]. This mapped expression data can be used to examine global and spatial expression patterns and to find genes with similar spatial expression profiles. Although the majority of Atlas transcripts represent protein coding genes, Mercer *et al.* [36] identified well over 1,000 Atlas riboprobes as putative lncRNAs and affirmed the expression patterns of some previously described lncRNAs, such as *Evf*, *Gtl2*, *Gomafu*, and *Sox2ot*.

Results

Structured Allen Mouse Brain Atlas riboprobes

Table 1 shows a summary of the Allen Mouse Brain Atlas riboprobes used in this study. For this study, we exclusively consider probes whose expression is above ISH background in the adult mouse brain [9]. Our main concern is whether "structured probes", *i.e.*, those containing a conserved RNA secondary structure as predicted by CMfinder, are in any other way a distinct population compared to "unstructured probes", *i.e.*, those lacking such predictions. Structured probes are further categorized into (1) putative ncRNAs (or simply ncRNAs) and (2) UTRs. A probe is classified as an ncRNA (of which some presumably are lncRNAs) if the entire probe is

Table 1 Structured Allen Mouse Brain Atlas riboprobes

	Only Intergenic	Only Intronic	Overlap 5' UTR	Overlap 3' UTR	Total
Expressed structured ncRNAs	141	10			151
Expressed structured UTRs			817	4,502	5,126
Expressed Atlas probes	462	15	2,492	8,108	16,483

Structured Allen Mouse Brain Atlas riboprobes have RNA secondary structures predicted by *CMfinder*, are expressed above background in the adult mouse brain and are mapped to the mouse genome (mm8). We subdivided structured probes into two genomic locations: (1) structured (long) noncoding RNAs (ncRNAs) have no protein coding potential and the entire probe is intergenic or intronic, and (2) structured UTRs are probes overlapping a UTR exon annotated in UCSC or RefSeq gene tracks with an RNA structure prediction. Total numbers of expressed Allen Mouse Brain Atlas probes (structured and non-structured) are listed. Detailed information about probes and structures are contained in Additional file 1.

intergenic or intronic without protein coding potential; it is classified as UTR if the probe overlaps an annotated UTR. Probes in coding exons are not analyzed.

By considering all annotated UTRs of full-length transcripts we find many additional structures, often at the end of longer alternative UTRs [37]. However, the expression of these variants in the brain is unknown, which is why we consider only those portions of UTRs that are overlapped by Atlas probes. For instance, one isoform of mouse VEGF-related factor gene (*Vegfb*) lacking a 3'UTR is expressed in brain [38]. We predict a RNA structure in one alternative 3'UTR of *Vegfb*, but we annotate it as a non-structured UTR because the Atlas probe does not overlap the extended 3' UTR structure.

CMfinder predicts conserved RNA structures in the genomic context of 11,998 Atlas riboprobes of which 10,516 probes are expressed above background (see Methods for mapping criteria). The amount of predicted structures overlapping expressed probes is sensitive to GC content but significantly larger in all GC bins than expected by chance (see Additional file 2: Table S1). We mapped the expressed structured probes to UCSC [39]

and RefSeq [40] gene tracks and obtained 5,126 probes with predicted structures in annotated untranslated regions (817 and 4,502 probes in 5' UTR and 3' UTR regions, respectively). The predicted structures are enriched at the flanks of UTRs, see Figure 1. In contrast, 4,467 expressed Atlas probes map to protein coding genes lacking structure predictions in their UTRs. Riboprobes not in annotated protein-coding genes are further examined for their protein-coding potential (see Methods for classification criteria). Excluding probes mapping to annotated coding exons and UTRs, we retain 141 intergenic and 10 intronic potential ncRNA transcripts that have predicted conserved local RNA structures (compared to 326 non-structured intergenic and intronic probes). Several RNA structures were found in known long ncRNAs such as *Xist*, *Miat*, *Meg3* and *Mirg*. Almost half (60) of the intergenic structured ncRNAs are more than 10 kb from the closest coding gene. Known RNA structures are annotated in nine putative ncRNA transcripts (4 microRNAs, 4 snoRNAs and *Xist*) and 80 structured UTRs (*e.g.*, 26 microRNAs, 19 snoRNAs, 6 SECIS, 6 Histone, and 3 IRES; see Additional file 2:

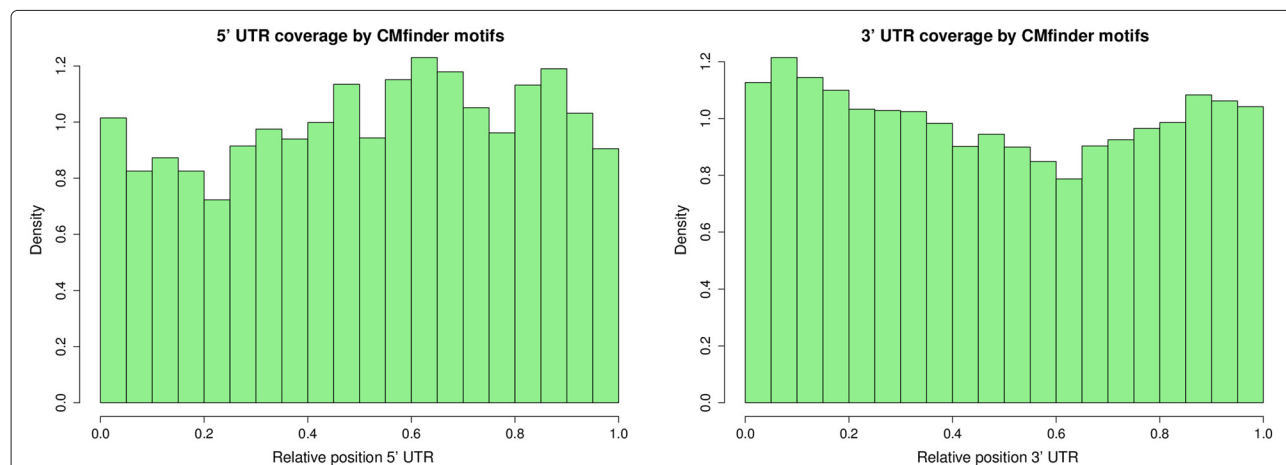


Figure 1 Relative location of predicted UTR structures. Relative location of *CMfinder* predicted structures in UTRs annotated in UCSC known genes that overlap the Allen Mouse Brain Atlas probes. Using *CMfinder* we predicted 1,367 structured loci (average length 76 nt) in 5' UTRs (average length 284 nt) of 817 expressed mRNAs and 11,551 structured loci (average length 82 nt) in 3' UTRs (average length 946 nt) of 4,502 expressed mRNAs. X-axis equal 0 describes the 5' end of UTRs.

Table S2). CMfinder predicts 27 of these annotated RNA structures, whereas the other predictions are located up- or downstream from known structures.

Spatial expression energy of structured transcripts

The Allen Mouse Brain Atlas has mapped the ISH expression to defined regions (see Additional file 2: Figure S1). To identify neuroanatomical-specific patterns of structured transcripts and to possibly gain some insight into the biological function of these transcripts, we apply a multi-resolution hierarchical search of increasing levels of granularity that starts with 11 neuroanatomical regions (cortex, CTX; olfactory bulb, OLF; hippocampus, HPF; striatum, STR; pallium, PAL; thalamus, TH; midbrain, MB; medulla, MY; hypothalamus, HY; cerebellum, CB; and pons, P) in sagittal sections and ends with three-dimensional grids of voxels each 200 micron per side for both the sagittal and coronal plane [41]. Unless stated otherwise, by structured we refer to a *predicted* RNA structure. First, we compare the mean expression level (technically, “expression energy”, as defined in [9]; see Methods) of expressed structured probes versus expressed non-structured probes. The comparisons were performed separately in each of the 11 neuroanatomical regions, and for both the putative ncRNA and UTR categories. Expression of all expressed Atlas probes was examined as well. For this analysis, we used sagittal tissue sections because sagittal data is available for all probes

whereas fewer (approximately 4,200 probes) have data in the coronal plane. In Figure 2, the major observations are that, first, structured UTR probes overall have the highest expression in the brain, and, second, structured ncRNA probes have higher expression than non-structured ncRNA probes, but are less strongly expressed than the average of all probes.

The significance of the observed patterns has been tested by different statistical methods (see Methods) with similar results. We addressed in which of the 11 neuroanatomical regions the mean expression energy of structured putative ncRNA and structured UTR probes is significantly different from expressed probes, intergenic and intronic, and UTR probes. The most striking result is that in all 11 neuroanatomical regions the 5,126 structured UTR probes have significantly higher expression than 4,467 non-structured UTR probes (see Additional file 2: Figure S2) as well as all expressed probes. The same applies for the finer level of granularity where the 11 neuroanatomical regions are further subdivided into 115 regions. On the other hand, there is significant expression enrichment of structured putative ncRNAs to non-structured ncRNAs only in cerebellum (using rmanovab).

Based on these observations we conducted further analyses to gain insight towards the possible causes of the enrichment of transcripts with structured UTRs. We studied significantly (p -value<0.001) enriched gene

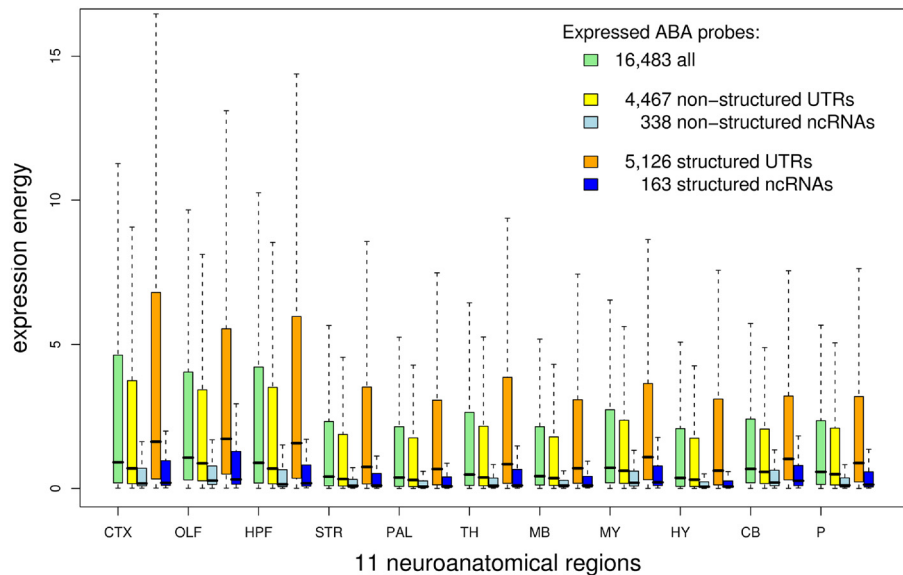


Figure 2 Expression energy distribution of the Allen Mouse Brain Atlas probes. Comparison of expression energy distribution of all expressed Allen Mouse Brain Atlas probes, structured and non-structured ncRNA and UTR probes in 11 neuroanatomical regions. Secondary RNA structures are predicted by CMfinder. The box plot shows 1.5 interquartile range (dotted line), lower and upper quartile (box), and median (thick black line in the box). Brain region abbreviations: cortex, CTX; olfactory bulb, OLF; hippocampus, HPF; striatum, STR; pallium, PAL; thalamus, TH; midbrain, MB; medulla, MY; hypothalamus, HY; cerebellum, CB; pons, P.

ontology (GO) terms [42] of UTR probes using functional annotation by DAVID [43]. We found support for function enrichment of binding ($p=5E-40$), localization ($p=4E-18$) and transport ($p=4E-16$) in structured UTR probes. Several GO terms for protein binding ($p=1E-43$) and RNA binding ($p=7E-7$) are significant for probes with structured UTRs (see Additional file 3), but none for non-structured UTRs (see Additional file 4). In addition, we found several GO terms which connect structured UTR probes to intracellular signaling pathways and suggest a directed RNA transport between nucleus and synapses or dendrites, e.g. the cellular components cytoplasm ($p=2E-35$), nucleus ($p=6E-15$) and synapse ($p=7E-6$) and the molecular functions intracellular signaling cascade ($p=2E-18$), protein transport ($p=4E-11$), protein localization ($p=1E-11$), vesicle-mediated transport ($p=1E-11$) and cytoskeletal protein binding ($p=1E-6$). For non-structured UTR probes there are four times less enriched GO terms, in general with lower significance than for structured UTR probes. Only the GO terms cytoplasm ($p=3E-22$) and transport ($p=5E-4$) are enriched for non-structured UTRs and are related to signaling function. Localization can imply different functional impact, for example direct involvement in transport, but it can also imply translational regulation at a specific subcellular location. Given the anatomy of the neurons where presumably many transcripts are located far away from the nucleus the observation of enriched expression of UTR regions with (predicted) RNA structures is consistent with this.

In embryonic cells it is known that the majority of localized RNAs are targeted to particular cytoplasmic regions by RNA elements and in mRNAs these are almost always in the 3' UTR [44]. In brain cells our data agrees with these earlier observations in the way that 5' UTR structures alone are not correlated to the binding, transporting and localization function of their protein products. We also mapped Atlas probes overlapping UTRs to a list of 76 active proteins in synapses [45]. A significantly greater number of their mRNAs has a structured UTR (Fisher's exact test p -value = 0.0023; see Additional file 2) which supports a role for the UTR structures as functional RNA elements. Another supporting observation for localized RNAs with structured RNA elements is their higher spatial divergence in the brain described by larger deviation of expression in 115 neuroanatomical regions (see Figure 3). Alternatively, spliced UTR transcripts may act independently from their host mRNA. However, this case cannot be verified without further examination of the probe captured transcripts.

The expression data also shows slightly higher expression of structured putative ncRNA transcripts than non-structured ncRNAs in many brain domains as indicated in Figure 2 and Additional file 2: Figure S3. Mean expression of the 151 structured ncRNA candidates is larger in 83

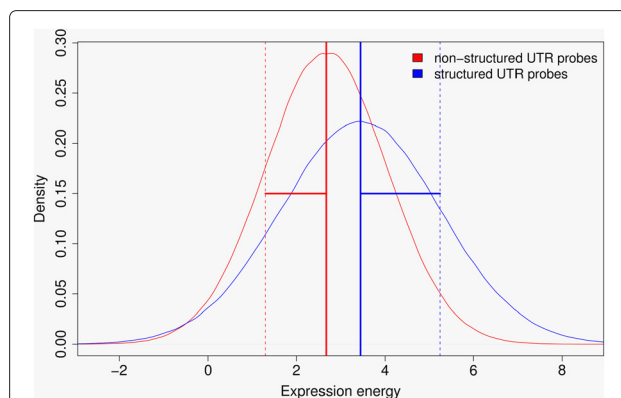


Figure 3 Spatial expression divergence. Mean normal distribution of expression energy of structured and non-structured UTR probes in 115 brain regions. The larger standard deviation (horizontal lines) of transcripts with structured UTRs shows their higher spatial expression divergence in the brain.

out of 115 brain regions. Enrichment is significant in 15 regions (including cerebellum) compared to 3 regions with significantly enriched non-structured ncRNA probes using a more robust measure of location assuming dependency between multiple ISH measurements (χ^2 manovab; see Methods).

It is essential to determine whether the presence of enriched transcripts is due to slower degradation caused by RNA structures [46]. The delay of degradation of transcripts folding in conserved RNA structures may support an increased half life of brain relevant RNAs. Proteins are actively synthesized in neuronal synapses despite the long distances between nucleus and synapses. For this purpose translational control of gene activity appears to be more efficient than transcriptional control [47]. Conservation of different structures in different transcripts suggests that they are involved in a rich variety of post-transcriptional regulatory interactions, e.g. through altered transcriptional stability. Combined with the previously described GO analyses, this suggest that proteins involved in molecule mobility are produced in larger numbers, and mRNAs and ncRNAs are transported to their intended cell destination before carrying out their function.

Protein binding of structured UTRs

As an initial step towards assigning functional information we searched for proteins that may bind to predicted structures in UTR regions. RNA binding proteins are trans-acting factors that function, e.g., in RNA localization. For instance, the mRNA of the neurotrophic tyrosine kinase *TrkB* receptor is transported to dendrites and translated in response to neural activity. The mouse *TrkB* 5' UTR contains one conserved and one mouse-specific single internal ribosomal entry site (IRES) whose RNA secondary structures and sequence-specific motifs

are proposed to be integral to IRES-dependent translation [48]. In agreement with this, the *CMfinder* prediction finds the conserved IRES structure in 9 mammals, whereas, as expected, the unconserved IRES structure was not predicted. The *CMfinder* structure consists of two stems of which the 3' stem is the same as previously shown in human [48]. Activity of the conserved IRES is enhanced in the presence of the polypyrimidine tract binding protein *PTBI* [49]. In the ISH data correlated expression of *TrkB* and *PTBI* can be seen, even though at a low level, in the olfactory bulb ($\rho = 0.49$) and medulla ($\rho = 0.52$) using the spatial homology search tool *NeuroBlast* [41] (see Methods).

In comparison to non-structured UTRs, a correlation-based search for similar expression pairs (using *NeuroBlast*) results in slightly more correlated expressed pairs between transcripts coding for RNA binding proteins and transcripts with structured UTRs. To identify spatial and brain-wide correlations, we used Pearson's correlation coefficient greater than a threshold of $\rho_T = 0.9$ and $\rho_T = 0.85$, respectively (see Methods for the selection criteria of ρ_T 's and spatial expression). We identified spatial correlation between 41 RNA binding proteins annotated in RBPDB [50] and 66 structured UTR transcripts mostly in thalamus, pallium and hippocampus (see Additional file 2: Table S3), as well as brain-wide correlated expression between 6 RNA binding proteins and 12 structured UTR probes (see Additional file 2: Table S4). We also searched for potential interaction sites of RNA binding proteins around UTR structures which are discussed below.

Correlated expression between structured transcripts

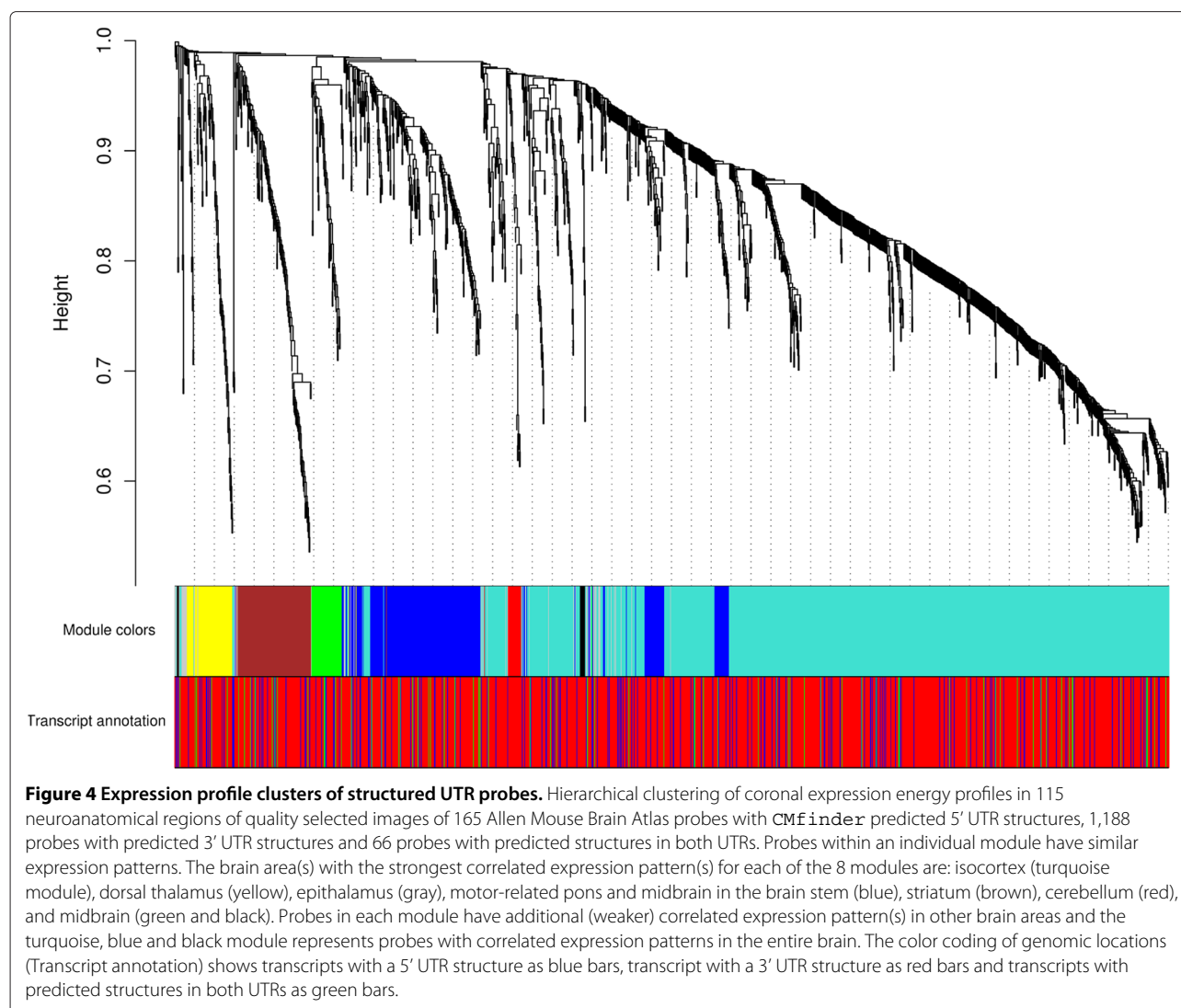
By examining correlated expression patterns, we can hypothesize new functions for previously uncharacterized structured transcripts or identify potential interacting RNA molecules as well as RNA-protein interactions due to localized translation as described above. The following *CMfinder* prediction of an annotated UTR element exemplifies connectivity of functional related molecules. We predict a widely conserved (in 16 organisms from human to zebrafish) 25 nt stem-loop in the 3' UTR of rat brain-derived neurotrophin factor *BDNF*. This stem-loop partly overlaps the loop and 5' end of the annotated core region of an extended stem-loop previously predicted in the full-length UTR structure (by *RNAfold*) [51]. The 3' UTR structure of *BDNF* provides a scaffold for interaction of various RNA binding proteins, polyadenylation factors and miRNAs in response to Ca^{2+} signal (neuron activity). The interaction results in Ca^{2+} signal-dependent stabilization of mRNAs in neurons [51].

Before studying gene pairs of correlated expression we look for groups of transcripts with structured UTRs with similar expression patterns in 115 brain regions.

High quality probes with coronal data (165 probes with structures in 5' UTRs, 1,188 probes in 3' UTRs and 66 probes in both UTRs; see Methods for selection criteria) are clustered in modules of correlated expression [52]. Most structured UTR probes have correlated expression patterns over the entire brain with the strongest signals in the isocortex (turquoise bar in Figure 4) and motor-related areas in the brain stem (blue bar). The strongest spatial pattern occurs in epithalamus (grey), followed by cerebellum (red), striatum (brown) and midbrain (green).

NeuroBlast is used to study correlated expression of structured Atlas probes in the entire brain. Starting with high quality probes we found 78 structured UTR transcripts with strong brain-wide expression involved in 352 brain-wide correlation pairs (threshold $\rho_T = 0.85$; see Additional file 2: Figure S4 for correlation network). Strong spatial activity is obtained for 264 structured UTR probes involved in 1,898 local correlation pairs ($\rho_T = 0.9$). Many transcripts have correlated expression to only a small number of other transcripts (see Additional file 2: Figure S5). One such example is the *Zfp365* zinc finger protein which is brain-wide correlated expressed to *6530418L21Rik* ($\rho = 0.86$), the signal transduction protein *Chn1* ($\rho = 0.87$) and *A230097P14Rik** ($\rho = 0.86$), whose mRNAs have highly conserved 3' UTR structures. Representative ISH images of some correlated probes are shown in Figure 5.

Sagittal image data was included for correlation pair analysis of ncRNAs. Of 477 putative ncRNAs, 9 show strong brain-wide correlated expression in 33 correlation pairs ($\rho_T = 0.8$) including 4 structured ncRNA candidates (*mCG145872*, *A230057G18Rik*, *TC1462951* and *Raph1*; see Additional file 2: Table S6 for a list of all correlated pairs). Most of these transcripts are involved in small cliques of correlated expression, see Figure 6. Additional file 2: Figure S6 shows representative ISH images of the non-coding myocardial infarction associated transcript *A230057G18Rik* (*Miat*) and its correlated expressed transcripts. More often ncRNA correlated expression appears in restricted brain domains rather than brain-wide. Mostly small cliques of correlated expressed transcripts are found for 134 ncRNAs (326 correlation pairs) including 33 structured ncRNA candidates involved in 84 correlation pairs ($\rho_T = 0.9$; see Additional file 2: Table S7 and Figure S7). Spatial correlation patterns also exist for known microRNAs and snoRNAs targeted by intergenic riboprobes. For instance, *mir-101a*, which is encompassed by AK021368 (*E130102H24Rik*), has correlated expression in hindbrain (*Emg1*) and pons, *mir-154* and *mir-410* which are encompassed by *Mirg* are expressed brain-wide with regional covariance in pons (*Mtch1*) and hippocampus (*Mrpl13*), and the *ACA17* snoRNA hosting transcript *mCG1030139* is correlated to *Mtnr1b* in the thalamus.



Thermodynamic stable RNA-RNA interactions

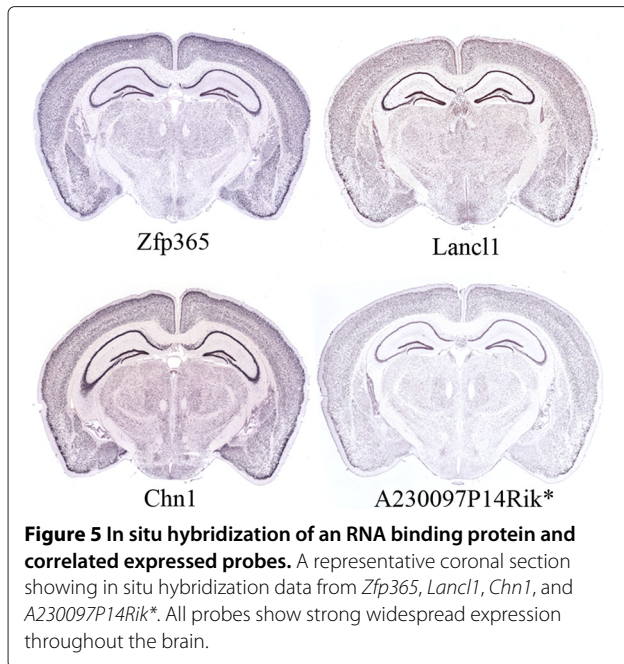
Many known ncRNAs exhibit their functionality through binding of RNA target sequences, such as microRNAs bind mRNAs, snoRNAs bind ribosomal and small nuclear RNAs, and certain lncRNAs may bind microRNAs [53] to regulate their activity or guide RNA editing. Potential RNA-RNA interactions between structured transcripts and correlated expressed RNAs were searched by scanning all putative ncRNAs and UTRs of Atlas transcripts for statistically significant intermolecular RNA binding sites. By combining *RNAplfold* [54] and *RNAplex* [55] we calculate the minimum free energy (MFE) of putative interaction sites in the real data, and the same strategy was used to create background distributions on dinucleotide shuffled data for *p*-value calculation (see Methods).

For 6 putative ncRNAs with local and 2 putative ncRNAs with brain-wide correlated expression we found putative interaction sites to 3' or 5' UTR of the correlated mRNAs, however, of relatively large *p*-values (see

Additional file 2: Table S8). For instance, a non-conserved interaction site is predicted between the putative ncRNA *TC1462951* and the 3' UTR of *Kcnb1* (see Additional file 2: Figure S8 for ISH image and expression mask). The putative ncRNA *LOC433503* may interact with a conserved region in the 3' UTR of *Gpx3*, only 100 nt upstream of the common stem-loop structure *SECIS* (see Additional file 2: Figure S9). In addition, around 600 significant (*p*-value < 1e-05) interaction sites with a MFE smaller than -40 kcal/mol are predicted by *RNAplfold* and *RNAplex* between structured putative ncRNAs and, e.g., UTRs of mRNAs coding for RNA binding proteins (*Rbpms* and *Samd14*; see Additional file 5), but the ISH data does not reveal correlated expression.

Discussion and Conclusions

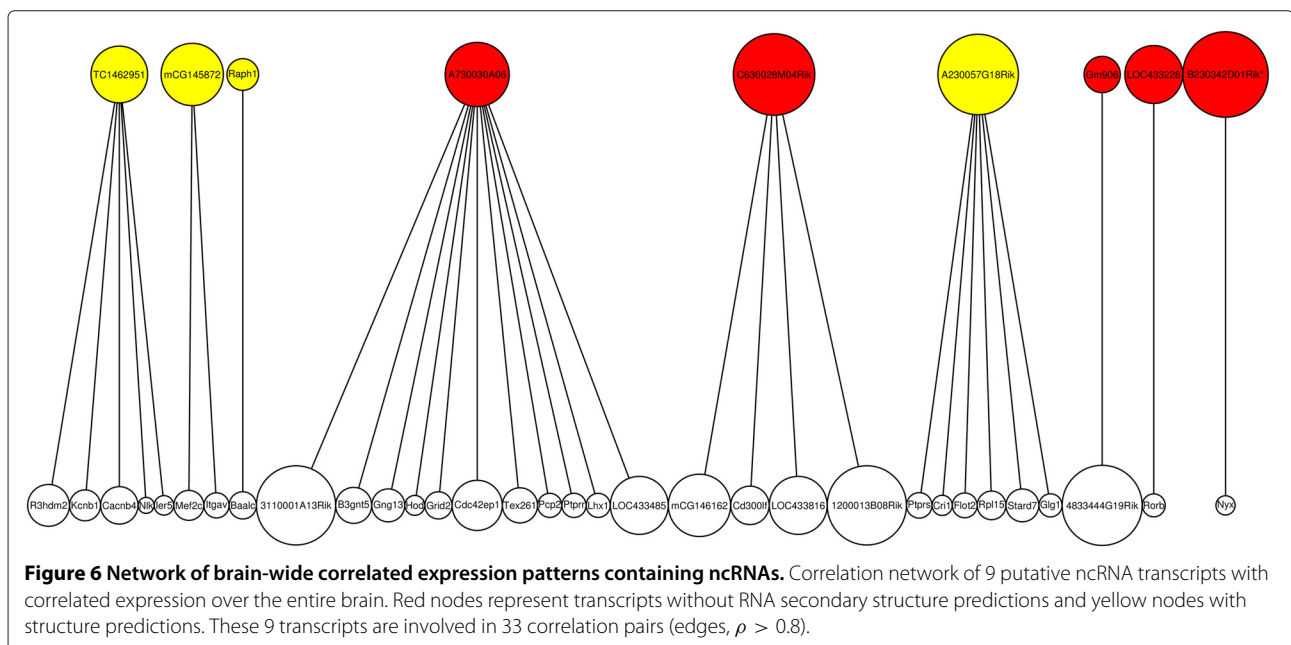
Microarray studies have shown that at least 50% of assayed transcripts are expressed in the brain [56], with up to 80% of transcripts shown to be expressed by ISH [9].



In order to gain a better understanding of transcripts in the brain that may be contributing to brain function, we examined which transcripts have an RNA structure. We observed that in silico predicted RNA structures are enriched both in coding (UTR regions) as well as noncoding transcripts in almost all regions of the adult mouse brain. The simplest interpretation of the data is that the Atlas probes showing higher expression are enriched for CMfinder predicted RNA structures. Through the integration of mouse brain expression data and secondary

RNA structure predictions, we found that transcripts with such predictions in their UTRs, those that are enriched in the 3' UTR adjacent to the ORE, have the highest expression throughout the brain. Many of these mRNAs as well as their protein products may act as signaling molecules whereas the UTR structures serve as binding motifs for other RNAs and proteins involved in intracellular signaling pathways. This hypothesis is supported by (i) enriched gene ontology terms binding, transport and localization, (ii) correlated expression patterns between mRNAs with structured UTRs and RNA binding proteins, and (iii) a larger expression diversity of transcripts with structured UTRs. UTR structures as signal for motor-driven transport and translational repression through RNA binding proteins are especially attractive in neurons where the transport of information stored in ribonucleic sequences from the nucleus through long axons to the synapses is an important component of neuronal functionality [47].

We investigated this hypothesis further by searching for potential protein binding motifs around (predicted) UTR structures to 72 RNA binding proteins annotated in RBPDB [50] (see Methods and Additional file 2). The majority (90%) of the UTR structures has at least one predicted binding motif in its neighborhood (see Additional file 2: Table S5). These motifs can be bound by 21 proteins. Only 9 proteins, however, have significantly more predicted targets than expected by chance, and half of the binding proteins are involved in splice site regulation. The analysis indicates that some interesting binding motifs can be found, such as neural-specific *Elavl2*, cytokine's degrading *Zfp36*, and mRNA trafficking *Khsrp*. *Zfp36* binds AU-rich elements (ARE) in the 3' UTR of



some cytokine mRNAs and promotes their degradation. Intriguingly, an AU-rich region (AU content of 85% over a length of 41 nt) starts at the 3' end of the predicted UTR structure of *6530418L21Rik* (see Additional file 2: Figure S10) and its expression is highly correlated with that of another zinc finger protein (*Zfp365*) and *Lancl1*, an RNA binding protein involved in immune surveillance of the brain [57] (see Figure 5). Assuming that *6530418L21Rik* works as a signaling molecule, its transport function may be deactivated through the binding of *Zfp36* close to its 3' UTR structure. However, here a large scale investigation in RNA-protein binding is still limited due to the low information content of binding motifs described by short sequence-based position weight matrices (PWMs).

Motivated by the GO analysis we also considered the hypothesis that structured RNAs in neural cells are themselves involved in establishing intracellular signaling pathways. For instance, Dienstbier *et al.* [3] provide evidence that Egalitarian (*EGL*) and the dynein cofactor Bicaudal D (*BICD*), previously known to be required for minus-end-directed mRNA transport, mediate linkage of various mRNAs to the dynein motor in *Drosophila melanogaster*. Here, we show that *EGL nine homolog 1* and *BICD* have predicted UTR structures, *BICD* is associated with the GO terms intramolecular, cytoplasm, localization, transport and binding and *EGL* with the GO term binding. Proteins, such as *EGL*, *BICD* and cytoskeletal protein filaments, are needed to establish intracellular pathways for directed cytoplasmic RNA transport towards synapses and dendrites. For signal propagation in the opposite direction back to the nucleus, mRNAs coding for these proteins have to be transported first and, thus, need cis-acting RNA elements too. The hypothesized directed RNA transport is illustrated in Additional file 2: Figure S14.

We also looked for predicted RNA structures in all UCSC and RefSeq annotated UTRs of protein coding genes overlapped by Atlas probes. We found 9,378 of these genes with RNA structure predictions in their UTRs and 5,576 without UTR structures. Of the 4,467 Atlas probes that overlap unstructured UTRs, 1,246 probes have a structure elsewhere in (at least one variant of) the UTR. It is unknown whether these structures are present in brain. Assuming they are, i.e., reclassifying as "structured" some of the UTR probes previously classified as "unstructured", we see even larger differences between the expression of structured and non-structured UTR probes (see Additional file 2: Figure S13 compared to Figure 2). Hence, we conclude that our overall statistics also hold for RNA structure annotation in full-length transcripts. In addition, we showed that putative ncRNAs with locally predicted RNA structures have significantly higher expression than non-structured intergenic and intronic transcripts in several brain regions.

Positive correlated expression patterns between pairs of transcripts are often domain-specific for putative structured ncRNAs. Most promising are 4 ncRNAs with brain-wide correlated expression in small cliques (*mCG145872*, *A230057G18Rik*, *TC1462951*, and *Raph1*; see Additional file 2: Table S6), and several ncRNAs with only one spatially correlated expressed transcript. We investigated conditions where RNA structure has a function, such as RNA-RNA interactions between correlated expressed RNA transcripts. One of the applied methods in this study, *e.g.* *RNAplex*, predicts the interaction site of two sequences. However, it is known from RNA motif searches that short sequence motifs can often appear by chance which partly explains the large *p*-values for the predicted RNA-RNA interactions. Consideration of homologous sequences in other species and duplex folding by using tools such as *PETcofold* [58,59] may help to obtain more significant predictions.

A major uncertainty is the limited resolution of the informatics detection of expression in the ISH images and, thus, the *NeuroBlast* correlation data. Several cells comprise a single voxel leading to interpolation between expression information and noisy expression energy. Sagittal images are more impacted by registration errors since only a single hemisphere is available for registration. The majority of correlation pairs detected in the sagittal plane failed validation by manual inspection of the ISH images (see Methods for further information). The largest cliques of correlated expression are often because of process artifacts in the images or the absence of expression (see Additional file 2: Figure S7). One desirable quality improvement of the correlation data is the weighted consideration of the voxel neighborhood which would improve the confidence in correlated expressed pairs by sacrificing some level of detail. Furthermore, the *NeuroBlast* data might also be interesting for graph theoretical analyses on gene expression correlation networks. Features of these networks are relatively unknown and the correlation coefficient threshold could be more sophisticatedly chosen by analyzing its influence on network connectivity. The large number of 3' UTR probes might also target ncRNAs, in addition to the untranslated region of mRNAs. In several specific cases we observed highly correlated brain-wide expression, *e.g.*, between the 3' UTR probe *Kcnc2* and its intronic *mCG142089*, and between *Dusp3* and its downstream-sense located structured probe *TC1462951*, but these probe pairs may have bound the same (pre-spliced) transcript. Thus, conclusions about correlated expression of adjacent or overlapping transcripts are hardly possible, especially if they have widespread expression throughout the brain.

An additional concern is that the observed correlation between structure and expression level might be an artifact of RNA degradation. All exonucleases have problems

initiating degradation close to stable stem structures [60]. Hence, the abundant enrichment of transcripts hosting RNA structures may be at least partly explained by their slower degradation and, thus, higher accessibility to riboprobes compared to transcripts lacking RNA structure. In fact, if the structures are involved in translational regulation, reduced degradation is just as effective as increased transcription in terms of raising steady-state transcript levels. Thus, to determine when *e.g.*, a bound protein primarily serves to regulate or primarily serves to prevent degradation seems hard, in particular if preventing degradation is part of the regulatory mechanisms as is the case with the iron metabolism in vertebrates [61]. However, the observed enrichment of transcripts with structured UTRs is not related to a particular structure, hence, it is unlikely that a particular RNA binding protein that promotes transcript stability by binding to a specific structured RNA motif is responsible for the broad expression pattern.

A final concern is that our results might be explained by a difference in the hybridization efficiency of Atlas probes towards structured versus unstructured transcripts. Hybridization is affected by a variety of factors, such as probe accessibility and affinity to the targeted molecule. For short oligos, although there are some contexts in which hybridization may be enhanced by appropriate RNA structures [62], it is most often suggested that highly structured regions in a target transcript would reduce hybridization efficiency. Many riboswitches, for example, down-regulate translation by sequestering the ribosome binding site in a structure that blocks interaction with the 16S rRNA [63]. This evidence suggests that structured target molecules would generate a decreased signal, but we observed an increase. In addition, Atlas probes were chosen to be 400-1200 bases in length. For such long probes that are perfectly complementary to their targets, the fully hybridized “double helix” will be the most energetically favorable state and seems likely to form easily from a simple initial toe-hold/zipper extension interaction from almost any initial conformation of the target. Thus, on balance, it does not seem likely that riboprobe affinity to structured versus unstructured transcripts explains the observed enrichment of structured transcripts.

Overall, our results show a huge potential for RNA structure as an abundant and active feature on both coding and noncoding transcripts in the adult mouse brain. Using *CMfinder* we predicted more than 40,000 RNA structures (mostly in intronic and 3'-untranslated regions) in about 10,500 expressed Atlas probes in the adult mouse brain. Even though *in silico* methods for RNA structure prediction hold high false positive rates of up to 50% [33,34] our findings still leave room for functional RNA structures in the Atlas transcriptome data. The significantly enriched expression energy of structured

transcripts is hard to explain by chance and supports the theme of functional RNA structures in the mouse brain. In the future, a structure analysis remains to be carried out on a global transcriptome data set in the adult mouse brain because the Atlas data primarily focus on protein-coding transcripts and has limited data on noncoding transcripts.

Methods

Mapping and classification criteria

The Allen Mouse Brain Atlas (Atlas) probes have been previously mapped to the mouse (mm8) genome [36]. Probe coordinates and RNA structure predictions are mapped to UCSC [39] and RefSeq [40] gene tracks with at least 10% overlap of probes and predictions. Intergenic and intronic probes are further checked for significant protein-coding potential as performed by Mercer *et al.* [36]: CRITICA [64] predicts significant protein-coding potential in the probe sequence or any targeted transcript, and ORFs greater than 120 codons are detected that comprise at least one third of the transcript length. In addition, we applied *RNAcode* [65] on mm8 based UCSC multiz17way alignments of intergenic and intronic probes to also detect shorter conserved ORFs (p -value < 0.001).

RNA structure predictions are in general unclear about which strand actually contains the structure [34]. Therefore, strand predictions of RNA structures were not used. We assume that a prediction on one strand yields a candidate on both strands. We mapped *CMfinder* structures to Atlas probes if the structure overlaps at least 1nt of an intergenic probe or if the structure overlaps at least 1nt of a UTR exon, coding exon, or intron that was mapped to the Atlas probe. We used this rather conservative procedure instead of mapping to putative respective transcripts of the probes to avoid counting splicing variants with predicted RNA structures. This procedure will miss some structured UTRs, however, our statistical conclusions still hold for the investigated subset of UTR structures.

Known RNA structures

The Allen Mouse Brain Atlas probes are annotated as known structured RNAs if they overlap at least 10% of a mouse microRNA in miRBase v10.0 [66] or a human track in miRBase, snoRNABase [67], Rfam 9.1 [68], ncRNA.org or Jones' and Eddy's ncRNA list [69]. We used *CMfinder* generated alignments and chained blastz alignments (liftOver tool) to map the human tracks to its mouse homologs.

Allen Mouse Brain Atlas technical information

The expression energy quantifies the overall expression at a given voxel. It is calculated as the product of expression level and density of cells expressed in that voxel [41]. All riboprobes have sagittal expression data and a subset

of riboprobes have both sagittal and coronal expression data. Informatics processing of the expression data from the sagittal sectioning plane is, however, effected by the data only containing one hemisphere (coronal data has two hemispheres), various starting and ending positions of the tissue sections processed for an individual riboprobe, and minor variability in the section cutting angle. In contrast, coronal data typically registers better as the symmetry of the section helps to lock the other two dimensions of the 3D grid together. To increase the accuracy of expression profiles and to meet quality control metrics, we created a high quality dataset that includes 1,525 structured UTR probes from Table 1 with coronal image series minus 125 coronal images series having manual detected processing artifacts (such as upside down images), widespread expression or missing image data due to failure of individual tissue sections.

Robust statistics

Significant spatial expression patterns are found by two sample location t-tests of the null hypothesis that the expression energy means of two sets of Atlas probes are equal. Errors associated with each ISH measurement are not totally independent from each other, thus, the normal distribution assumption does not hold. We apply bootstrap procedures to estimate the unknown distribution of expression energy in neuroanatomical regions. The percentile-t bootstrap p -values differ from ordinary percentile p -values in that they are based on bootstrap approximations of the distribution of the studentized estimator rather than the distribution of the original estimator. P -values are adjusted by the method of Benjamini & Hochberg to control the false discovery rate and the null hypothesis was rejected if the adjusted p -value < 0.25. As a more robust measurement of location we also calculated adjusted p -values of 0.2% trimmed means using the bootstrap methods `rmmcp` and `rmanovab` [70].

NeuroBlast and Pearson's correlation coefficient

The Atlas provides interpolated expression energy in regular 3-dimensional lattices of cellular resolution for each sagittal and coronal image series. The correlation of the expression energy for each probe pair is calculated by the spatial homology search tool `NeuroBlast` [41]. `NeuroBlast` calculates the Pearson's correlation coefficient ρ between two vectors of two probes that hold the expression energies for all voxels each 200 micron per side in a defined brain region. The cumulative frequency distribution of the number of correlation pairs over ρ follows typically a negative sigmoid curve (see Additional file 2: Figure S11), thus we chose a threshold ρ_T close to the right flattened area of the curve for selecting the most promising correlation pairs. Spatial

correlations have tendential higher ρ 's than brain-wide correlations due to the lower amount of compared voxels. Hence, we chose ρ_T slightly higher to select spatial correlations.

Spatial expression

Riboprobes with high spatial expression are defined as probes with larger relative expression in one brain domain compared to the entire brain:

$$\frac{v_d(E \geq 2)}{v_d} / \frac{v_b(E \geq 2)}{v_b} > 1,$$

where v_d is the number of voxels in one domain and v_b is the number of voxels in the entire brain.

Protein binding sites

UTR structures and their 50 nt flanking regions are searched for potential protein binding motifs using RBPDB [50]. First, position weight matrices (PWMs) from RBPDB were used together with the perl TFBS library [71] to scan sequences for binding sites to 72 RNA binding proteins with expressed Atlas probes (461 proteins in RBPDB). Second, we sequence aligned (BLAT) our sequences against 1,021 individual RNA sequences from single-sequence experiments excluding consensus (IUPAC) sequences.

Prediction of significant RNA-RNA interactions

Potential interaction sites of all ncRNAs included in the Atlas are searched in all UTRs of Atlas transcripts annotated in mouse by UCSC or RefSeq. Probabilities of local basepairs are calculated by `RNAplfold` in all sequences. These probabilities are taken as input for `RNAplex` for considering sequence accessibility. `RNAplex` v0.2 is used with the parameters `-c 40 -e -10 -l 30 -z 30`. We calculate a p -score for putative interaction sites which is the probability of obtaining a MFE S by chance greater than the observed MFE. Therefore, we dinucleotide shuffled 100 times all queries and targets of the top 10,000 interaction pairs and calculated their binding MFE. Additional file 2: Figure S12 shows that the MFEs are extreme-value distributed (evd) with a maximum around -10 (which is used as censored cutoff for evd parameter estimation). Since MFE is highly dependent on the length and GC content of the interaction site, we describe the background distribution (λ and μ) for 49 combinations of the two covariates using the `HMMER` package [72] and estimate a p -value of predicted RNA-RNA interactions on the appropriate background by the equation:

$$P(S \geq \text{MFE}) = 1 - \exp(-e^{\lambda(\text{MFE}-\mu)})$$

Additional files

Additional file 1: Annotation of predicted structured probes. CSV-file listing the features of all structured probes from Table 1 and their *CMfinder* predicted RNA secondary structures.

Additional file 2: Tables and figures. This file contains lists of correlated expressed structured riboprobes, and additional tables and figures.

Additional file 3: GO analysis of structured UTR probes. 4115 structured UTR probes with known gene symbols are examined for GO term enrichment.

Additional file 4: GO analysis of non-structured UTR probes. 3407 non-structured UTR probes with known gene symbols are examined for GO term enrichment.

Additional file 5: Predicted significant RNA-RNA interactions. CSV-file listing 585 significant (p -value < 1e-05) interactions between structured putative ncRNAs and UTRs. The interaction sites are predicted by *RNAfold* and *RNAplex* to be larger than 9 nt and with a MFE smaller than -40 kcal/mol.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We thank Quaid Morris for the helpful discussion about RNA binding proteins and Marcel Dinger for the probe mapping to the mouse genome and the coding potential pipeline. SMS and MJH thank the Allen Institute for Brain Science founders, Paul G. Allen and Jody Allen, for their vision, encouragement, and support. SES and JG were supported by the Lundbeck Foundation, the Danish Council for Independent Research (Technology and Production Sciences), the Danish Council for Strategic Research (Programme Commission on Strategic Growth Technologies), as well as the Danish Center for Scientific Computing. SMS and MJH were supported by the Allen Institute for Brain Science.

Author details

¹Center for non-coding RNA in Technology and Health, University of Copenhagen, Frederiksberg C, Denmark. ²IBHV, University of Copenhagen, Frederiksberg C, Denmark. ³Allen Institute for Brain Science, Seattle, USA. ⁴Computer Science and Engineering and Genome Sciences, Univ. of Washington, Seattle, USA. ⁵Fred Hutchinson Cancer Research Center, Seattle, USA.

Authors' contributions

SES did the analyses and wrote the main part of the manuscript. SMS and MJH helped in the integration of the Allen Mouse Brain Atlas expression data, and SMS manually inspected the ISH images. JG and WLR had the idea for the study and helped to integrate the RNA structure predictions. All authors contributed to the manuscript.

Received: 9 September 2011 Accepted: 31 May 2012

Published: 31 May 2012

References

- Jansen R: **mRNA localization: message on the move.** *Nat Rev Mol Cell Biol* 2001, **2**(4):247–256.
- Czaplinski K, Singer R: **Pathways for mRNA localization in the cytoplasm.** *Trends Biochem Sci* 2006, **31**(12):687–693.
- Dienstbier M, Boehl F, Li X, Bullock S: **Egalitarian is a selective RNA-binding protein linking mRNA localization signals to the dynein motor.** *Genes Dev* 2009, **23**(13):1546–1558.
- Kubota S, Mukudai Y, Moritani N, Nakao K, Kawata K, Takigawa M: **Translational repression by the cis-acting element of structure-anchored repression (CAESAR) of human ctgf/ccn2 mRNA.** *FEBS Lett* 2005, **579**(17):3751–3758.
- Gorodkin J, Cirera S, Hedegaard J, Gilchrist MJ, Panitz F, Jørgensen C, Scheibye-Knudsen K, Arvin T, Lumholdt S, Sawera M, Green T, Nielsen BJ, Havgaard JH, Rosenkilde C, Wang J, Li H, Li R, Liu B, Hu S, Dong W, Li W, Yu J, Wang J, Staefeldt HH, Wernersson R, Madsen LB, Thomsen B, Hornshøj H, Bujie Z, Wang X, Wang X, Bolund L, Brunak S, Yang H, Bendixen C, Fredholm M: **Porcine transcriptome analysis based on 97 non-normalized cDNA libraries and assembly of 1,021,891 expressed sequence tags.** *Genome Biol* 2007, **8**(4):R45.
- Seemann SE, Gilchrist MJ, Hofacker IL, Stadler PF, Gorodkin J: **Detection of RNA structures in porcine EST data and related mammals.** *BMC Genomics* 2007, **8**:316 [http://www.ncbi.nlm.nih.gov/pubmed/17845718].
- Mercer TR, Dinger ME, Mariani J, Kosik KS, Mehler MF, Mattick JS: **Noncoding RNAs in Long-Term Memory Formation.** *Neuroscientist* 2008, **14**(5):434–445 [http://www.ncbi.nlm.nih.gov/pubmed/18997122].
- Ponjavic J, Oliver PL, Lunter G, Ponting CP: **Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain.** *PLoS Genet* 2009, **5**(8):e1000617.
- Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen L, Chen TM, Chin MC, Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, Desaki AL, Desta T, Diep E, Dolbeare TA, Donelan MJ, Dong HW, Dougherty JG, Duncan BJ, Ebbert AJ, Eichele G, Estin LK, Faber C, Facer BA, Fields R, Fischer SR, Fliss TP, Frensley C, Gates SN, Glattfelder KJ, Halverson KR, Hart MR, Hohmann JG, Howell MP, Jeung DP, Johnson RA, Karr PT, Kawai R, Kidney JM, Knapik RH, Kuan CL, Lake JH, Laramee AR, Larsen KD, Lau C, Lemon TA, Liang AJ, Liu Y, Luong LT, Michaels J, Morgan JJ, Morgan RJ, Mortrud MT, Mosqueda NF, Ng LL, Ng R, Orta GJ, Overly CC, Pak TH, Parry SE, Pathak SD, Pearson OC, Puchalski RB, Riley ZL, Rockett HR, Rowland SA, Royall JJ, Ruiz MJ, Sarno NR, Schaffnit K, Shapovalova NV, Sivisay T, Slaughterbeck CR, Smith SC, Smith KA, Smith BI, Sodt AJ, Stewart NN, Stumpf KR, Sunkin SM, Sutram M, Tam A, Teemer CD, Thaller C, Thompson CL, Varnam LR, Visel A, Whitlock RM, Wohnoutka PE, Wolkey CK, Wong VY, Wood M, Yaylaoglu MB, Young RC, Youngstrom BL, Yuan XF, Zhang B, Zwingman TA, Jones AR: **Genome-wide atlas of gene expression in the adult mouse brain.** *Nature* 2007, **445**(7124):168–176 [http://www.ncbi.nlm.nih.gov/pubmed/17151600].
- Babendure JR, Babendure JL, Ding JH, Tsien RY: **Control of mammalian translation by mRNA structure near caps.** *RNA* 2006, **12**(5):851–861.
- Nackley A, Shabalina S, Tchivileva I, Satterfield K, Korchymskiy O, Makarov S, Maixner W, Diatchenko L: **Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure.** *Science* 2006, **314**(5807):1930–1933.
- Kloc M, Zearfoss NR, Etkin LD: **Mechanisms of subcellular mRNA localization.** *Cell* 2002, **108**(4):533–544.
- Yaniv K, Yisraeli JK: **Defining cis-acting elements and trans-acting factors in RNA localization.** *Int Rev Cytol* 2001, **203**:521–539.
- Chartrand P, Singer RH, Long RM: **RNP localization and transport in yeast.** *Annu Rev Cell Dev Biol* 2001, **17**:297–310.
- Aranda-Abreu G, Hernandez M, Soto A, Manzo J: **Possible Cis-acting signal that could be involved in the localization of different mRNAs in neuronal axons.** *Theor Biol Med Model* 2005, **2**:33.
- Gonzalez I, Buonomo S, Nasmyth K, von Ahnen U: **ASH1 mRNA localization in yeast involves multiple secondary structural elements and Ash1 protein translation.** *Curr Biol* 1999, **9**(6):337–340.
- Wedemeyer N, Schmitt-John T, Evers D, Thiel C, Eberhard D, Jockusch H: **Conservation of the 3'-untranslated region of the Rab1a gene in amniote vertebrates: exceptional structure in marsupials and possible role for posttranscriptional regulation.** *FEBS Lett* 2000, **477**(1-2):49–54.
- Ainger K, Avossa D, Diana A, Barry C, Barbaresi E, Carson J: **Transport and localization elements in myelin basic protein mRNA.** *J Cell Biol* 1997, **138**(5):1077–1087.
- Ryder S, Williamson J: **Specificity of the STAR/GSG domain protein Qk1: implications for the regulation of myelination.** *RNA* 2004, **10**(9):1449–1458.
- Carninci P: **Tagging mammalian transcription complexity.** *Trends Genet* 2006, **22**(9):501–510.
- Mattick JS: **The genetic signatures of noncoding RNAs.** *PLoS Genet* 2009, **5**(4):e1000459.
- Mercer TR, Dinger ME, Mattick JS: **Long non-coding RNAs: insights into functions.** *Nat Rev Genet* 2009, **10**(3):155–159.
- Ponting CP, Oliver PL, Reik W: **Evolution and functions of long noncoding RNAs.** *Cell* 2009, **136**(4):629–641.

24. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES: **Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.** *Nature* 2009, **458**(7235): 223–227.
25. Ng JH, Ng HH: **LincRNAs join the pluripotency alliance.** *Nat Genet* 2010, **42**(12):1035–1036.
26. Lipovich L, Johnson R, Lin CY: **MacroRNA underdogs in a microRNA world: Evolutionary, regulatory, and biomedical significance of mammalian long non-protein-coding RNA.** *Biochim Biophys Acta* 2010, **1799**(9):597–615.
27. Zhou Y, Zhong Y, Wang Y, Zhang X, Batista DL, Gejman R, Ansell PJ, Zhao J, Weng C, Klibanski A: **Activation of p53 by MEG3 non-coding RNA.** *J Biol Chem* 2007, **282**(34):24731–24742.
28. Huarte M, Guttman M, Feldser D, Garber M, Koziol MJ, Kenzelmann-Broz D, Khalil AM, Zuk O, Amit I, Rabani M, Attardi LD, Regev A, Lander ES, Jacks T, Rinn JL: **A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response.** *Cell* 2010, **142**(3): 409–419.
29. Pedersen JS, Bejerano G, Siepel A, Rosenbloom K, Lindblad-Toh K, Lander ES, Kent J, Miller W, Haussler D: **Identification and Classification of Conserved RNA Secondary Structures in the Human Genome.** *PLoS Comput Biol* 2006, **2**(4):e33.
30. Washietl S, Hofacker IL, Stadler PF: **Fast and reliable prediction of noncoding RNAs.** *Proc Natl Acad Sci USA* 2005, **102**(7):2454–9.
31. Yao Z, Weinberg Z, Ruzzo WL: **CMfinder – a covariance model based RNA motif finding algorithm.** *Bioinformatics* 2006, **22**(4):445–52.
32. Torarinsson E, Sawera M, Havgaard JH, Fredholm M, Gorodkin J: **Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure.** *Genome Res* 2006, **16**(7):885–9.
33. Gorodkin J, Hofacker IL, Torarinsson E, Yao Z, Havgaard JH, Ruzzo WL: **De novo prediction of structured RNAs from genomic sequences.** *Trends Biotechnol* 2010, **28**:9–19.
34. Gorodkin J, Hofacker I: **From Structure Prediction to Genomic Screens for Novel Non-Coding RNAs.** *PLoS Comput Biol* 2011, **7**(8):e1002100.
35. Ng L, Bernard A, Lau C, Overly CC, Dong HW, Kuan C, Pathak S, Sunkin SM, Dang C, Bohland JW, Bokil H, Mitra PP, Puellas L, Hohmann J, Anderson DJ, Lein ES, Jones AR, Hawrylycz M: **An anatomic gene expression atlas of the adult mouse brain.** *Nat Neurosci* 2009, **12**(3):356–362.
36. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS: **Specific expression of long noncoding RNAs in the mouse brain.** *Proc Natl Acad Sci USA* 2008, **105**(2):716–21 [http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=18184812].
37. Nagasaki H, Arita M, Nishizawa T, Suwa M, Gotoh O: **Species-specific variation of alternative splicing and transcriptional initiation in six eukaryotes.** *Gene* 2005, **364**:53–62.
38. Townson S, Lagercrantz J, Grimmond S, Silins G, Nordenskjold M, Weber G, Hayward N: **Characterization of the murine VEGF-related factor gene.** *Biochem Biophys Res Commun* 1996, **220**(3):922–928.
39. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human genome browser at UCSC.** *Genome Res* 2002, **12**(6):996–1006.
40. Pruitt KD, Maglott DR: **RefSeq and LocusLink: NCBI gene-centered resources.** *Nucleic Acids Research* 2001, **29**:137–40.
41. Ng L, Pathak SD, Kuan C, Lau C, Dong H, Sodt A, Dang C, Avants B, Yushkevich P, Gee JC, Haynor D, Lein E, Jones A, Hawrylycz M: **Neuroinformatics for genome-wide 3D gene expression mapping in the mouse brain.** *IEEE/ACM Trans Comput Biol Bioinform* 2007, **4**(3):382–393.
42. Ashburner M, Ball C, Blake J, Botstein D, Butler H, Cherry J, Davis A, Dolinski K, Dwight S, Eppig J, Harris M, Hill D, Issel-Tarver L, Kasarskis A, Lewis S, Matese J, Richardson J, Ringwald M, Rubin G, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25–29.
43. Huang DW, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**:44–57.
44. Bashirullah A, Cooperstock R, Lipshitz H: **RNA localization in development.** *Annu Rev Biochem* 1998, **67**:335–394.
45. Bayes A, van de Lagemaat L, Collins M, Croning M, Whittle I, Choudhary J, Grant S: **Characterization of the proteome, diseases and evolution of the human postsynaptic density.** *Nat Neurosci* 2011, **14**:19–21.
46. Houseley J, Tollervey D: **The many pathways of RNA degradation.** *Cell* 2009, **136**(4):763–776.
47. Presutti C, Rosati J, Vincenti S, Nasi S: **Non coding RNA and brain.** *BMC Neurosci* 2006, **7**(Suppl 1):S5.
48. Dobson T, Minic A, Nielsen K, Amiot E, Krushel L: **Internal initiation of translation of the TrkB mRNA is mediated by multiple regions within the 5' leader.** *Nucleic Acids Res* 2005, **33**(9):2929–2941.
49. Timmerman S, Pflingsten J, Kieft J, Krushel L: **The 5' leader of the mRNA encoding the mouse neurotrophin receptor TrkB contains two internal ribosomal entry sites that are differentially regulated.** *PLoS One* 2008, **3**(9):e3242.
50. Cook KB, Kazan H, Zuberi K, Morris Q, Hughes TR: **RBPDB: a database of RNA-binding specificities.** *Nucleic Acids Res* 2010.
51. Fukuchi M, Tsuda M: **Involvement of the 3'-untranslated region of the brain-derived neurotrophic factor gene in activity-dependent mRNA stabilization.** *J Neurochem* 2010, **115**(5):1222–1233.
52. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC Bioinformatics* 2008, **9**:559.
53. Wilusz JE, Sunwoo H, Spector DL: **Long noncoding RNAs: functional surprises from the RNA world.** *Genes Dev* 2009, **23**(13):1494–1504.
54. Bernhart SH, Hofacker IL, Stadler PF: **Local RNA base pairing probabilities in large sequences.** *Bioinformatics* 2006, **22**(5):614–615.
55. Tafer H, Hofacker IL: **RNAplex: a fast tool for RNA-RNA interaction search.** *Bioinformatics* 2008, **24**(22):2657–2663.
56. Sandberg R, Yasuda R, Pankratz DG, Carter TA, Del Rio JA, Wodicka L, Mayford M, Lockhart DJ, Barlow C: **Regional and strain-specific gene expression mapping in the adult mouse brain.** *Proc Natl Acad Sci U S A* 2000, **97**(20):11038–11043.
57. Mayer H, Bauer H, Breuss J, Ziegler S, Prohaska R: **Characterization of rat LANCL1, a novel member of the lanthionine synthetase C-like protein family, highly expressed in testis and brain.** *Gene* 2001, **269**(1–2):73–80.
58. Seemann S, Richter A, Gesell T, Backofen R, Gorodkin J: **PETcofold: predicting conserved interactions and structures of two multiple alignments of RNA sequences.** *Bioinformatics* 2011, **27**(2):211–219.
59. Seemann S, Menzel P, Backofen R, Gorodkin J: **The PETfold and PETcofold web servers for intra- and intermolecular structures of multiple RNA sequences.** *Nucleic Acids Res* 2011, **39**(Web Server issue):W107–11.
60. Deutscher MP: **Degradation of RNA in bacteria: comparison of mRNA and stable RNA.** *Nucleic Acids Res* 2006, **34**(2):659–666.
61. Walden W, Selezneva A, Dupuy J, Volbeda A, Fontecilla-Camps J, Theil E, Volz K: **Structure of dual function iron regulatory protein 1 complexed with ferritin IRE-RNA.** *Science* 2006, **314**(5807): 1903–1908.
62. Mir K, Southern E: **Determining the influence of structure on hybridization using oligonucleotide arrays.** *Nat Biotechnol* 1999, **17**(8):788–792.
63. Meyer M, Ames T, Smith D, Weinberg Z, Schwalbach M, Giovannoni S, Breaker R: **Identification of candidate structured RNAs in the marine organism 'Candidatus Pelagibacter ubique'.** *BMC Genomics* 2009, **10**:268.
64. Badger JH, Olsen GJ: **CRITICA: coding region identification tool invoking comparative analysis.** *Mol Biol Evol* 1999, **16**(4):512–524.
65. Washietl S, Findeiss S, Muller S, Kalkhof S, von Bergen M, Hofacker I, Stadler P, Goldman N: **RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data.** *RNA* 2011, **17**(4):578–594.
66. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ: **miRBase: tools for microRNA genomics.** *Nucleic Acids Research* 2008, **36**(Database issue):D154–D158 [http://nar.oxfordjournals.org/cgi/content/full/36/suppl_1/D154].
67. Lestrade L, Weber MJ: **snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs.** *Nucleic Acids Res* 2006, **34**(Database issue):D158–D162.
68. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A: **Rfam:**

- updates to the RNA families database.** *Nucleic Acids Res* 2009, **37**(Database issue):D136–D140.
69. Jones TA, Eddy SR: **ncRNA annotation track for the human genome, version hg16 (July 2003).** 2004.
 70. Wilcox RR(Ed): *Introduction to robust estimation and hypothesis testing*: Elsevier Academic Press; 2005.
 71. Lenhard B, Wasserman WW: **TFBS: Computational framework for transcription factor binding site analysis.** *Bioinformatics* 2002, **18**(8):1135–1136.
 72. Eddy SR: **Maximum likelihood fitting of extreme value distributions.**, 1997. [<http://selab.janelia.org/publications/Eddy97b/Eddy97b-techreport.ps>] [Technical Report, Janelia Farm].

doi:10.1186/1471-2164-13-214

Cite this article as: Seemann *et al.*: Transcripts with *in silico* predicted RNA structure are enriched everywhere in the mouse brain. *BMC Genomics* 2012 **13**:214.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

